Appl Microbiol Biotechnol (1986) 23:355–360

Applied and Microbiology Biotechnology

Comparison of amino acid sequences of eleven different α-amylases

Ryoichi Nakajima, Tadayuki Imanaka, and Shuichi Aiba

Department of Fermentation Technology, Faculty of Engineering, Osaka University, Yamada-oka, Suita-shi, Osaka 565, Japan

Summary. A comparison was made of the amino acid sequences of 11 different α-amylases. The 6 animal α-amylases tested were found to be highly homologous (about 80 to 90%, depending on the species compared). Amino acid sequence of *Bacillus stearothermophilus* α-amylase was fairly homologous (about 60%) with that of a thermostable α-amylase from *Bacillus amyloliquefaciens*. Homology was least among the thermolabile amylases from *Bacillus subtilis, Aspergillus oryzae,* plants and animals. Nevertheless, four highly homologous regions were found in the amino acid sequences of all the enzymes, despite their widely different origins. It was inferred that these four homologous regions were likely to be the active and/or substrate-binding sites.

Introduction

α-Amylase (1,4-α-D-glucan glucanohydrolase; EC 3.2.1.1) is a widely distributed secretory enzyme. The enzyme has been investigated extensively from various aspects: its protein structure and function, its mechanism of secretion through cell membrane, and its industrial application. Several workers have already published amino acid sequences of α-amylase, either deduced from the nucleotide sequences of the cloned genes (Hagenbüchle et al. 1980; MacDonald et al. 1980; Rogers and Milliman 1983; Takkinen et al. 1983; Yamazaki et al. 1983; Yang et al. 1983; Nakamura et al. 1984; Nakajima et al. 1985) or by direct determination by amino-acid analysis (Kluh 1981; Toda et al. 1982).

Offprint requests to: T. Imanaka

The amino acid sequences and structural properties of a large number of different proteins are to be found in the extensive literature on the subject (Kyte and Doolittle 1982). If this information could be made available, it would in principle be possible to account for the difference in heat resistance shown by the same enzyme from different origins; that is, from organisms that are thermophiles, mesophiles or psychrophiles. Whatever the principle might be, the feature which determines the thermophilic nature of a protein has not yet been made clear. Some reports suggest that thermostability of a protein can be enhanced by a single amino-acid substitution (Yutani et al. 1977; Matsumura et al. 1984).

The purpose of this analysis was to align the known amino acid sequences of α-amylases from different sources in an attempt to detect any regions common to all the enzymes, however, different their origins.

Procedures

Alignment of amino acid sequences. The amino acid sequences of α-amylases shown in Fig. 1 were deduced from a determination of nucleotide sequences from several publications (Hagenbüchle et al. 1980; MacDonald et al. 1980; Rogers and Milliman 1983; Takkinen et al. 1983; Yang et al. 1983; Nakamura et al. 1984; Nakajima et al. 1985). This does not apply to amino acid sequences of hog pancreatic α-amylase and Taka-amylase of *Aspergillus oryzae:* which were determined directly from their cyanogen bromide digests after preparing crystalline samples (Kluh 1981; Toda et al. 1982).

The sources of α-amylase are given on the left of the figure and the sequences are read from left to right. Except for barley, the signal peptide regions are placed in the 1st row of each sequence. Each row for the same enzyme origin is composed, unless otherwise noted, of 60 amino acids from left to right, and the last amino acid is joined sequentially by the first one in the next row.

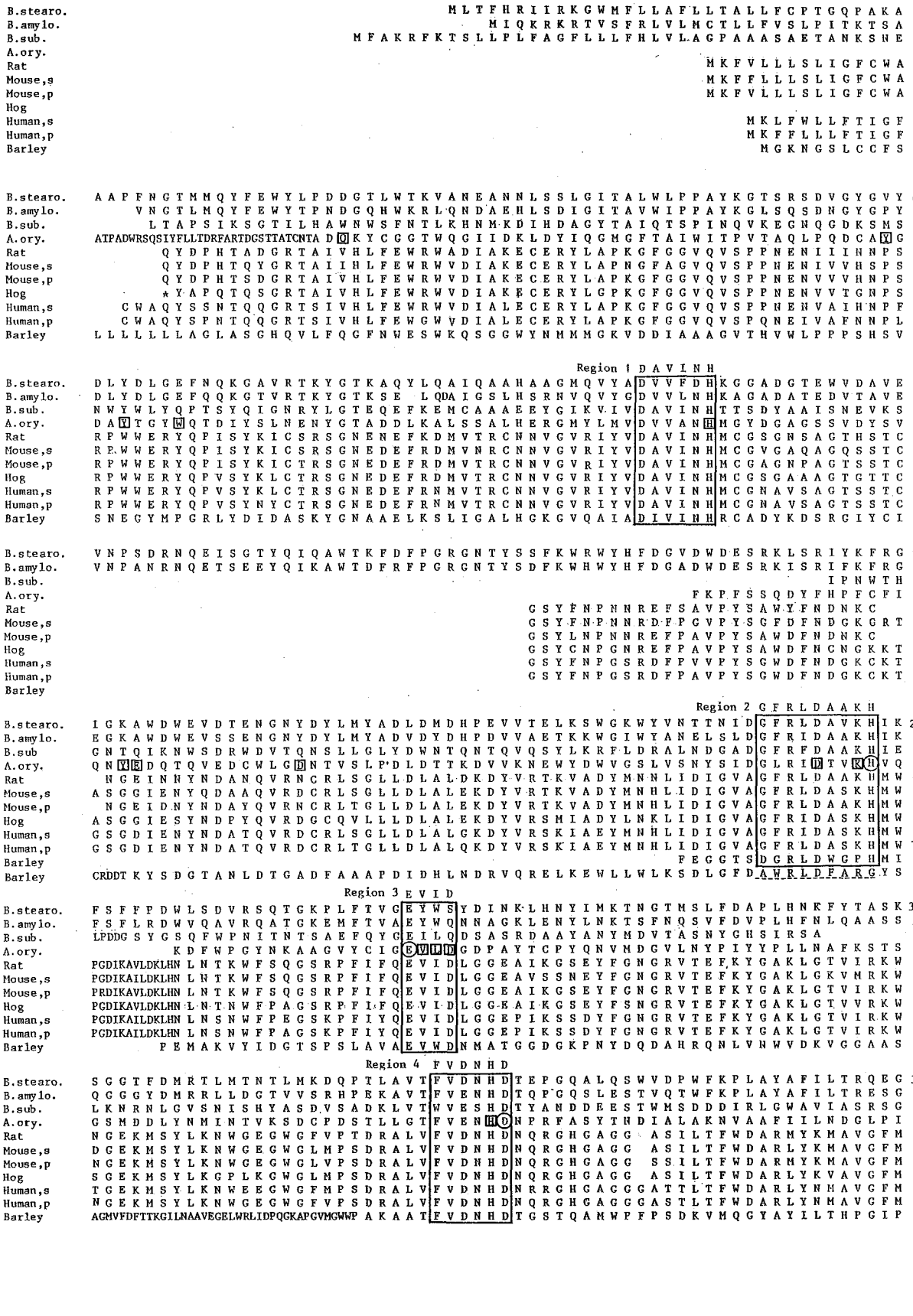
B.stearo.
B.amylo.
B.sub.
A.ory.
Rat
Mouse,s
Mouse,p
Hog
Human,s
Human,p
Barley
Region 1 D A V I N H
Region 2 G F R L D A A K H
Region 3 E V I D
Region 4 F V D N H D

Fig. 1. Comparison of amino acid sequences of various α-amylases. Amino acid residues are shown by single letters as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Enzyme sources are abbreviated as: B. stearo., *Bacillus stearothermophilus*; B. amylo., *Bacillus amyloliquefaciens*; B. sub., *Bacillus subtilis*; A. ory., *Aspergillus oryzae*; s, saliva; p, pancreas. The first amino acid of extracellular amylase of *B. stearothermophilus*, Ala, is counted as +1. Signal peptides are shown in the first rows. Homologous sequence regions 1, 2, 3 and 4 are surrounded by rectangles. Amino acid sequence described above the rectangular regions was taken as representative of regions 1 to 4, respectively. Active sites and those of substrate binding proposed by Matsuura et al. (1984) for Taka-amylase A from *A. oryzae* are indicated by ○, and □, respectively. Barley amylase has another sequence somewhat similar to region 2 as shown by a broken underline. *, Pyrrolid-2-one-5-carboxylic acid (Kluh 1981)

Some anomalies of alignment are noted in Fig. 1. The interval of arrangement between the neighbouring amino acids is halved and/or expanded, whenever required, to the extent of considerable or total absence of amino acids in some rows of this alignment (see the dented area of the left-hand side of the figure). These anomalies arose either from a trial and error or from a computer-aided search for sequential regions common to all of the enzymes studied (regions 1 to 4, designated by rectangular areas in Fig. 1; for details of the location of these regions, see later). Thus the absence of some amino acids from rows of the same origin does not have any special significance; it has resulted principally from an attempt to facilitate the comparison.

With regard to α-amylase from barley which is composed of 438 amino acid residues, a converse alignment was attempted, after having fixed their regions 1 to 4 (loc. cit.). Accordingly, 11 amino acid residues appearing in the 1st row have nothing to do with the signal sequence in this case. The double rows for barley in the 5th segment from the top likewise are of no special significance: this duplicate row was introduced to extract a segment (with a broken line beneath it) that was homologous with the corresponding regions of other sources (region 2 in Fig. 1). Similarly, the arrangement of triple rows for *B. subtilis* in the last segment of this arrangement was necessary because this α-amylase, composed of 660 amino acid residues, is characterized by a very long sequence in the COOH-terminal region. It is reported that at least 101 amino acid residues of the COOH-terminus could be removed without the loss of the enzyme activity (Yamazaki et al. 1983).

Finally, it should be mentioned that signal sequences of mouse salivary and pancreatic α-amylases shown in Fig. 1 are in the signal sequence for the rat enzyme (MacDonald et al. 1980; MacDonald et al. 1982); Hagenbüchle et al. (1980) have described the signal sequence, comprising the first 12 amino acid residues, for mouse enzyme. By the same token, signal sequences of the human salivary and pancreatic α-amylases might well be altered in the same way; this has not been done, however, and we use the sequence shown in the original paper (Nakamura et al. 1984). However inconsistent this treatment might appear, the arrangement as in Fig. 1 would not have any serious effect on the conclusions to be drawn from this comparison of 11 different α-amylases, or on the identification of homologous regions.

Analysis of the primary-structure homology and the hydropathic character of α-amylase. The search for homologous regions in amino acid sequences was aided partly by an NEC PC-8001 computer (Nippon Electric Co., Tokyo, Japan), using a program of the dot matrix (Novotny 1982). The hydropathic character of α-amylase of *B. stearothermophilus* (taken as an example) was also analysed by the computer as described earlier (Matsumura et al. 1984).

Comparison and discussion

It should be pointed out that, as can be seen in Fig. 1, the signal sequences of the genus *Bacillus* are long (29 to 41 amino acid residues), while those of the enzyme of higher animals are relatively short (12 to 15 amino acid residues); the signal sequence of amylase from barley has, however, not been determined (Rogers and Milliman 1983).

Homology of amino acid residues between two α-amylases is defined as the ratio of the number of identical amino acid residues to the total number of residues of the smaller sequence. The homology was 80—90% in 6 different animal α-amylases. Homology among the α-amylases from a wider range of living organisms — microorganisms, plants and animals, was no more than 10%; an exception was the 60% homology between *B. stearothermophilus* and *B. amyloliquefaciens* α-amylases.

It seemed necessary to resort to the use of a computer to search for homologous regions among the different α-amylases. A dot matrix plot for *B. amyloliquefaciens* and *B. stearothermophilus* α-amylases is shown in Fig. 2A, that for α-amylases of *B. subtilis* and *B. stearothermophilus* is shown in Fig. 2B. A clear, diagonal dotted line in Fig. 2A demonstrates a high degree of homology between *B. amyloliquefaciens and B. stearothermophilus* (thermostable) α-amylases. There is, however, no such distinct diagonal line of dotted points in Fig. 2B; close examination, based on the evidence in Fig. 1, showed that regions 2 and 4 could be recognized, although not very clearly. This reflects the low degree of homology between *B. subtilis* (thermolabile) and *B. stearothermophilus* (thermostable) α-amylases.

Since these two regions were found to be shared by all the amylases examined, as is seen in Fig. 1, it was concluded that regions 2 and 4 might function as active and/or substrate-binding sites.

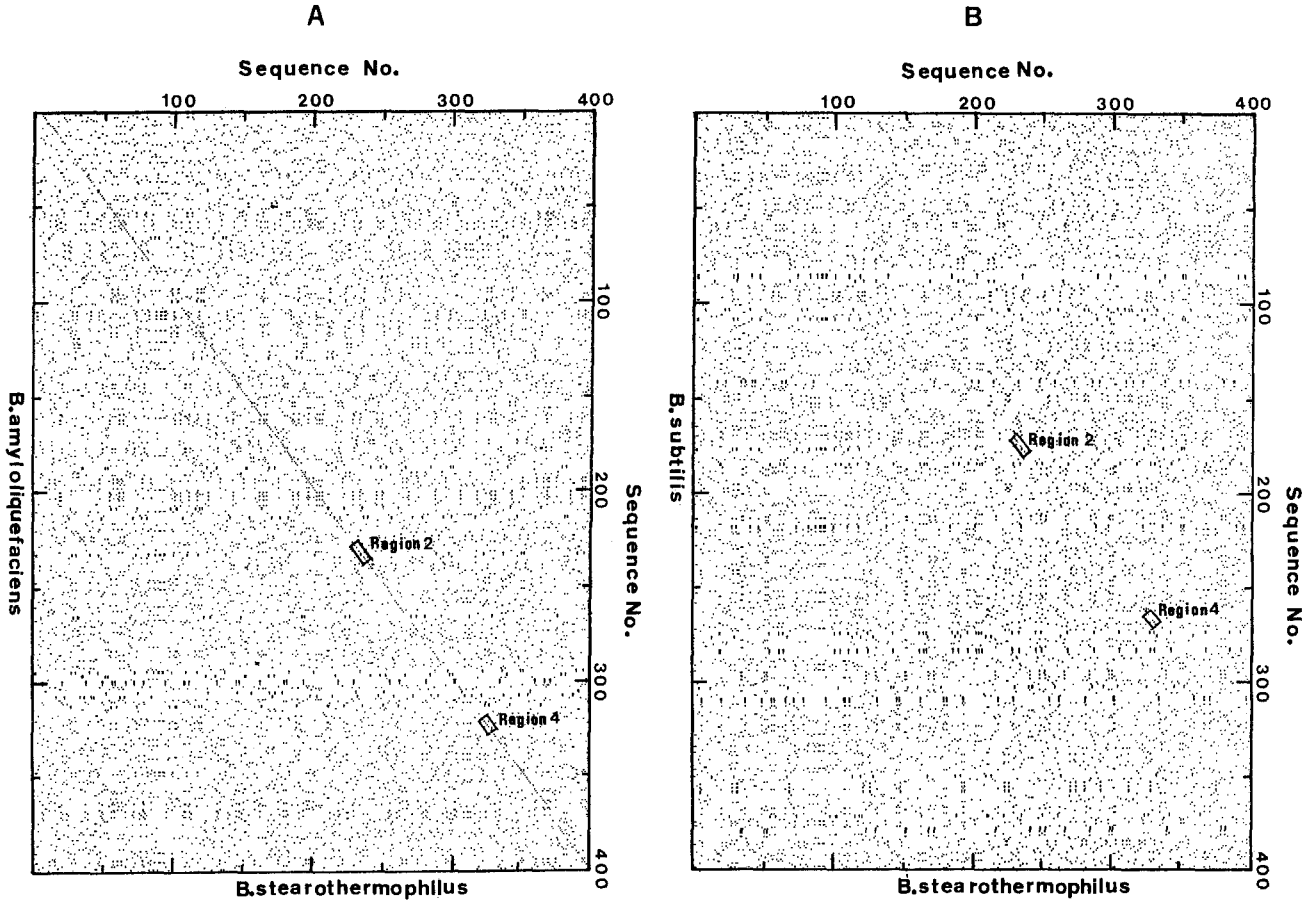

Fig. 2. Computer search for homologous area of the amino acid sequence. A) *B. amyloliquefaciens* vs *B. stearothermophilus*, B) *B. subtilis* vs *B. stearothermophilus*. Areas surrounded by boxes correspond to regions 2 and 4 in Fig. 1. The sequence numbers in both diagrams are for the secreted proteins, counting the 1st amino acid in the 2nd row in Fig. 1 as +1. Regions 1 and 3 are not shown in these diagrams: for these, see text

In fact, a molecular model of Taka-amylase A (α-amylase) from *A. oryzae* by Matsuura et al. (1984) suggests that both His and Asp in these homologous areas function as active sites, while Asp and Lys in region 2 and His in region 4 may function as substrate-binding sites.

Since another active site of Glu and the neighboring substrate-binding sites of Val, Leu and Asp between regions 2 and 4 have been suggested (Matsuura et al. 1984), the sequences were examined for an alignment containing these four amino acid residues. Just such an area has been designated as region 3 in Fig. 1. It is interesting to note that the three active-site amino-acids, His, Glu and Asp, occur consistently in regions 2, 3 and 4, respectively for all the enzymes tested. Many further substrate-binding sites are proposed for Taka-amylase A (Matsuura et al. 1984); these are indicated by the square symbols scattered throughout the entire sequence of amylase from *A. oryzae*.

In the process of locating homologous areas in the amino acid sequences by the trial and error procedure, we found, in addition to regions 2 and 4, another homologous area, designated as region 1 in Fig. 1. The fact that only these four homologous regions 1, 2, 3 and 4 were found consistently, despite the different origins of the organisms, whether bacteria, fungi, higher plants and animals, suggests that these regions are indispensable for the expression of α-amylase activity.

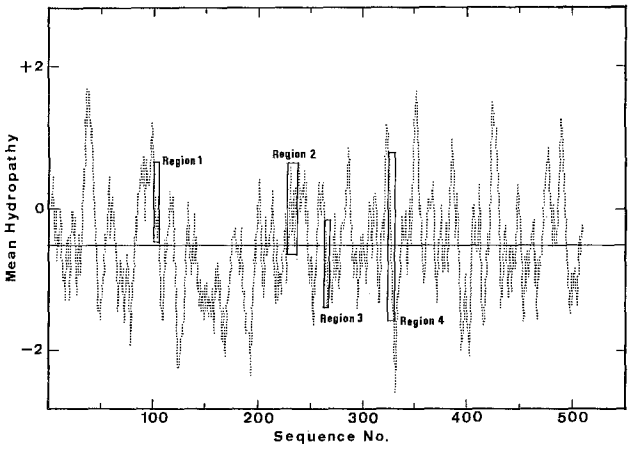

Fig. 3. Hydropathy profile of *B. stearothermophilus* α-amylase. The mean hydropathy (Kyte and Doolittle 1982) of a moving segment of nine amino acid residues is shown. The sequence No. is defined as before. The line parallel to the abscissa represents the average of hydropathy of this enzyme. The areas corresponding to regions 1, 2, 3 and 4 in Fig. 1 are surrounded by boxes

Figure 3 shows the hydropathic character of α-amylase from *B. stearothermophilus*. Hydrophobicity and hydrophilicity of the enzyme are determined by a moving segment of nine amino acid residues along the sequence. The parts that project above the solid line in the figure denote strongly hydrophobic and interior regions of the enzyme, whereas those that project below the line designate strongly hydrophilic segments and the exterior. Regions 1, 2, 3 and 4 reproduced in Fig. 3 are located in the valley of protein surface between the exterior and interior regions. This supports the suggestion that these regions are important for the catalytic activity of α-amylase.

Another characteristic difference between the thermostable enzymes of *B. stearothermophilus, B. amyloliquefaciens* and the thermolabile enzyme from *B. subtilis* is the distance between regions 1 and 2 (Fig. 1). These differences may indicate a possible line of research on the complex nature of thermophily.

References

Hagenbüchle O, Bovey R, Young RA (1980) Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isozyme mRNAs from pancreas and salivary gland. Cell 21:179—187

Kluh I (1981) Amino acid sequence of hog pancreatic α-amylase isoenzyme I. FEBS Lett 136:231—234

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105—132

MacDonald RJ, Crerar MM, Swain WF, Pictet RL, Thomas G, Rutter WJ (1980) Structure of a family of rat amylase genes. Nature 287:117—122

MacDonald RJ, Stary SJ, Swift GH (1982) Two similar but nonallelic rat pancreatic trypsinogens: nucleotide sequences of the cloned cDNAs. J Biol Chem 257:9724—9732

Matsumura M, Katakura Y, Imanaka T, Aiba S (1984) Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. J Bacteriol 160:413—420

Matsuura Y, Kusunoki M, Harada W, Kakudo M (1984) Structure and possible catalytic residues of Taka-amylase A. J Biochem 95:697—702

Nakajima R, Imanaka T, Aiba S (1985) Nucleotide sequence of the *Bacillus stearothermophilus* α-amylase gene. J Bacteriol 163:401—406

Nakamura Y, Ogawa M. Nishide T, Emi M, Kosaki G, Himeno S, Matsubara K (1984) Sequences of cDNAs for human salivary and pancreatic α-amylases. Gene 28:263—270

Novotny J (1982) Matrix program to analyze primary structure homology. Nucleic Acids Res 10:127—131

Rogers JC, Milliman C (1983) Isolation and sequence analysis of a barley α-amylase cDNA clone. J Biol Chem 258:8169—8174

Takkinen K, Pettersson RF, Kalkkinen N, Palva I, Söderlund H, Kääriäinen L (1983) Amino acid sequence of α-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. J Biol Chem 258:1007—1013

Toda H, Kondo K, Narita K (1982) The complete amino acid sequence of Taka-amylase A. Proc Japan Acad 58:208—212

Yamazaki H, Ohmura K, Nakayama A, Takeichi Y, Otozai K, Yamasaki M, Tamura G, Yamane K (1983) α-Amylase genes (*amyR2* and *amyE*$^+$) from an α-amylase-hyperproducing *Bacillus subtilis* strain: Molecular cloning and nucleotide sequences. J Bacteriol 156:327—337

Yang M, Galizzi A. Henner D (1983) Nucleotide sequence of the amylase gene from *Bacillus subtilis*. Nucleic Acids Res 11:237—249

Yutani K, Ogasahara K, Sugino Y, Matsushiro A (1977) Effect of a single amino acid substitution on stability of conformation of a protein. Nature 267:274—275

Received July 25, 1985/Revised September 12, 1985